NAME

rabins − split **argus(8)** data.

SYNOPSIS

rabins −M *splitmode* [*splitmode options*]] [**raoptions**]

DESCRIPTION

Rabins reads **argus** data from an *argus-data* source, and adjusts the data so that it is aligned to a set of bins, or slots. The resulting output is split, modified, and optionally aggregated so that the data fits to the constraints of the specified bins. **rabins** is designed to be a combination of **rasplit** and **racluster**.

The principal function of rabins is to align data on a time series. This is critical for real-time stream block processing, graphing, comparing, analyzing, and correlating argus data. But rabins() also supports bin'ing for a fixed number of argus data records ('count'), or a fixed volume of data ('size'). While these two last options are very useful, they are rather esoteric. See the online examples and rasplit.1 for examples of using these modes of operation.

Time Series Bins

Time series bin'ing is specified using the **-M** *time* option. Time bins are specified by the size and granularity of the time bin. The granularity, such as 'd' for days, dictates where the bin boundaries lie (in this case, the boundary starts at the beginning of each day).

Programatically, each bin gets its own complete racluster.1 function,

Records that span a time boundary are split, so that the data represents that fraction that resides in the bin, with the metrics adjusted in a uniformly distributed fashion. The result is a series of data and/or fragments that are time aligned, and is appropriate for time seried analysis, and visualization.

When a record is split to conform to a time series bin, either the starting or ending timestamps can fall within a specified time boundary. In some applications, it is desired that the timestamps conform to the time bin boundaries, however in some applications having the exact times is critical to retain transaction duration and burst behavior. Rabins supports the optional **hard** option to specify that timestamps should conform to 'hard' boundaries, forcing **rabins** to modify the start and stop timestamps in records to the time series slot boundaries. One of the results of this is that all durations in the reported records will be the slot duration. This is extremely important when printing certain metrics, like average load.

The output files name consists of a prefix, which is specified using the *-w ra option*, and for all modes except **time** mode, a suffix, which is created for each resulting file. If no prefix is provided, then **rabins** will use 'x' as the default prefix. The suffix that is used is determined by the mode of operation. When **rabins** is using the default count mode or the size mode, the suffix is a group of letters 'aa', 'ab', and so on, such that concatenating the output files in sorted order by file name produces the original input file. If **rabins** will need to create more output files than are allowed by the default suffix strategy, more letters will be added, in order to accomodate the needed files.

When **rabins** is spliting based on time, **rabins** uses a default extension of %Y.%m.%d.%h.%m.%s. This default can be overrided by adding a '%' extension to the name provided using the *-w* option.

When standard out is specified, using *-w* -, **rabins** will output a single **argus-stream** with START and STOP argus management records inserted appropriately to indicate where the output is split. See **argus(8)** for more information on output stream formats.

When **rabins** is spliting on output record count (the default), the number of records is specified as an ordinal counter, the default is 1000 records. When **rabins** is spliting based on the maximum output file size, the size is specified as bytes. The scale of the bytes can be specified by appending 'b', 'k' and 'm' to the

number provided.

When **rabins** is spliting base on time, the time period is specified with the option, and can be any period based in seconds (s), minutes (m), hours (h), days (d), weeks (w), months (M) or years (y). **Rabins** will create and modify records as required to split on prescribed time boundaries. If any record spans a time boundary, the record is split and the metrics are adjusted using a uniform distribution model to distribute the statistics between the two records.

RABINS SPECIFIC OPTIONS

Rabins, like all ra based clients, supports a number of **ra options** including remote data access, reading from multiple files and filtering of input argus records through a terminating filter expression. **rabins(1)** specific options are:

−a *suffix length*
default is 2 characters.

−M *splitmode*
Supported spliting modes are:
time <period>
count <n[kmb]
size <n[kmb]>
soft
nomodify

−m *aggregation object*
Supported aggregation objects are:
none - use a null flow key.
srcid - argus source identifier.
smac - source mac(ether) addr.
dmac - destination mac(ether) addr.
smpls[ind] - source mpls label
dmpls[ind] - destination mpls label
svlan - source vlan label.
dvlan - destination vlan label.
saddr - source IP addr.
daddr - destination IP addr.
proto - transaction protocol.
sport - source sap.
dport - destination sap.
stos - source TOS byte value.
dtos - destination TOS value.
sttl - source TTL value.
dttl - destination TTL value.
stcpb - source TCP base seq number.
dtcpb - destination TCP base seq number.

−w *filename*
Rabins supports an extended *-w* option that allows for output record contents to be inserted into the output filename. Specified using '$' (dollar) notation, any printable field can be used. Care should be taken to honor any shell escape requirements when specifying on the command line. See **ra(1)** for the list of printable fields.

Another extended feature, when using **time** mode, **rabins** will process the supplied filename using **strftime(3)**, so that time fields can be inserted into the resulting output filename.

INVOCATION

This invocation reads **argus(8)** data from **inputfile** and splits the **argus(8)** data stream based on output file size of no greater than 1 Megabyte. The resulting output files have a prefix of *argus.* and suffix that starts with 'aa'.

rabins -r argusfile -M soft time 1m -s +1dur -m proto - ip

This invocation splits **inputfile** based on hard 10 minute time boundaries. The resulting output files are created with a prefix of */archive/%Y/%m/%d/argus.* and the suffixes *%H.%M.%S*. The values will be supplied based on the time in the record being written out.

rabins -r * -M time 10m -w "/archive/%Y/%m/%d/argus.%H.%M.%S"

This invocation splits **inputfile** based on the argus source identifier. The resulting output files are created with a prefix of */archive/Source Identifier/argus.* and the default suffix starting with "aa". The source identifier will be supplied based on the contents of the record being exported.

rabins -r * -M time 10m -w "/archive/\$srcid/argus."

COPYRIGHT

SEE ALSO

ra(1), rarc(5), argus(8),

AUTHORS

Carter Bullard (carter@qosient.com).